

BAB I PENDAHULUAN

1.1 Latar Belakang

Kecurangan dalam transaksi e-commerce telah menjadi masalah serius yang merugikan banyak pihak, terutama bagi pemilik bisnis dan konsumen. Berbagai bentuk kecurangan, seperti pencurian identitas, transaksi palsu, dan penyalahgunaan kartu kredit, dapat mengakibatkan kerugian finansial yang signifikan. Berdasarkan laporan, kecurangan dalam transaksi e-commerce dapat menyebabkan kerugian miliaran rupiah setiap tahunnya, menciptakan kebutuhan mendesak untuk menemukan solusi yang efektif guna mendeteksi dan mencegah kejadian tersebut. Tindakan-tindakan ini mempunyai dampak negatif terhadap bisnis dan klien, menimbulkan risiko serius terhadap keuangan, reputasi [1].

Penelitian sebelumnya telah banyak dilakukan dalam upaya mendeteksi kecurangan menggunakan berbagai algoritma machine learning. Namun, akurasi deteksi yang dihasilkan sering kali tidak optimal, terutama disebabkan oleh ketidakseimbangan data, di mana jumlah transaksi yang sah jauh lebih banyak dibandingkan dengan transaksi curang. Ketidakseimbangan kelas merupakan kondisi dimana jumlah data pada kelas minoritas lebih sedikit dibandingkan dengan kelas mayoritas. Dampak dari ketidakseimbangan kelas pada dataset adalah terjadinya kesalahan klasifikasi kelas minoritas, sehingga dapat mempengaruhi kinerja klasifikasi. [2] Terdapat metode yang sering digunakan dalam menyelesaikan masalah klasifikasi, antara lain Logistic Regression (LR), Support Vector Machine (SVM), Naive Bayes, XGBoost, Decision Tree dan Random Forest (RF). Namun penggunaan metode-metode ini untuk kasus fraud detection pada sektor perbankan memiliki tantangan tersendiri [3]. Namun, algoritma tersebut juga memiliki kelemahan, seperti overfitting, kesulitan menangani data yang sangat tidak seimbang, dan hasil prediksi yang kurang stabil tanpa penyetelan parameter yang tepat.

Beberapa menghasilkan model prediksi yang kurang optimal dan yang lain memerlukan penyesuaian nilai hyperparameternya [4]. Untuk mengatasi masalah ini, teknik penyetelan parameter atau hyperparameter tuning menjadi penting untuk meningkatkan kinerja algoritma dalam mendeteksi kecurangan. Dengan memilih parameter yang tepat, kelemahan seperti overfitting pada Random Forest atau ketidakmampuan Support Vector Machine untuk mengelola data besar secara efisien dapat diminimalkan. Dalam penelitian ini, metode machine learning dengan penyesuaian parameter digunakan untuk meningkatkan akurasi dan stabilitas model, sehingga dapat memberikan hasil yang lebih baik dibandingkan dengan pendekatan sebelumnya.



1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Seberapa efisien penerapan pendekatan metode SMOTE dalam mengatasi ketidakseimbangan data dalam meningkatkan kinerja ML pada kecurangan transaksi pada e-commerce?
2. Apakah penggunaan penyetelan hyperparameter menggunakan grid search dapat meningkatkan kinerja algoritma ML dalam mendeteksi kecurangan pada transaksi di e-commerce?
3. Algoritma mana yang memiliki kinerja terbaik dalam mendeteksi kecurangan pada transaksi di e-commerce ?



1.3 Batasan Masalah

Untuk menjaga fokus penelitian dan memastikan hasil yang lebih relevan, beberapa batasan dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya berfokus pada deteksi kecurangan transaksi pada e-commerce menggunakan data yang sudah tersedia, tidak mencakup transaksi di sektor lain seperti perbankan atau asuransi.
2. Algoritma machine learning yang digunakan dalam penelitian ini terbatas pada Logistic Regression, Random Forest, XGBoost, GaussianNB, Decision Tree, SVM tanpa membahas algoritma lain yang lebih kompleks.
3. Data yang digunakan hanya berupa dataset transaksi yang telah tersedia dan tidak mencakup data real-time atau data transaksi yang berasal dari sumber lain.
4. Penelitian ini hanya menggunakan teknik resampling (undersampling) untuk mengatasi masalah ketidakseimbangan data, tanpa membandingkan dengan teknik lain seperti oversampling atau SMOTE.
5. Hasil akurasi yang didapat dari model hanya dianalisis berdasarkan kinerja pada dataset yang diberikan, tanpa melakukan uji coba di lingkungan produksi atau e-commerce secara langsung.
6. Variabel yang dianalisis dalam penelitian ini terbatas pada variabel yang ada dalam dataset, sementara variabel lain yang mungkin relevan dengan kecurangan transaksi namun tidak tersedia dalam dataset akan diabaikan. batasan masalah penelitian yang dilakukan.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menghasilkan model machine learning yang mampu mendeteksi transaksi kecurangan pada e-commerce dengan tingkat akurasi yang optimal.
2. Menganalisis dan membandingkan kinerja beberapa algoritma machine learning, seperti Logistic Regression, Random Forest, dan XGBoost, dalam mendeteksi kecurangan pada transaksi e-commerce.
3. Mengatasi masalah ketidakseimbangan data dalam dataset transaksi e-commerce melalui teknik resampling, guna meningkatkan kinerja model dalam mendeteksi kecurangan.
4. Memvalidasi model yang dihasilkan dengan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score, untuk memastikan efektivitas model dalam mendeteksi kecurangan.
5. Memberikan rekomendasi mengenai algoritma machine learning yang paling efektif digunakan dalam mendeteksi transaksi yang mencurigakan pada platform e-commerce.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik dari segi teoritis maupun praktis, yang dapat dijelaskan sebagai berikut:

1. Manfaat Teoritis:

- o Menambah wawasan dan pengetahuan dalam bidang penerapan machine learning untuk deteksi kecurangan transaksi pada e-commerce.
- o Mengembangkan referensi akademik terkait pemanfaatan algoritma machine learning seperti Logistic Regression, Random Forest, dan XGBoost dalam menangani masalah kecurangan transaksi.
- o Memberikan kontribusi terhadap literatur yang membahas teknik resampling untuk mengatasi ketidakseimbangan data pada dataset yang digunakan dalam deteksi kecurangan.

2. Manfaat Praktis:

- o Bagi **platform e-commerce**, hasil penelitian ini dapat digunakan sebagai landasan untuk mengimplementasikan sistem deteksi dini terhadap transaksi yang mencurigakan, sehingga mengurangi risiko kerugian akibat kecurangan.
- o Bagi **pengembang teknologi dan praktisi**, penelitian ini memberikan gambaran yang jelas mengenai model dan algoritma machine learning yang paling efektif dalam mendeteksi kecurangan, yang dapat diadopsi dan dikembangkan lebih lanjut.
- o Bagi **peneliti selanjutnya**, penelitian ini dapat menjadi dasar untuk mengembangkan model prediksi kecurangan yang lebih akurat atau untuk mengaplikasikan algoritma machine learning lain yang belum dibahas, serta memperluas penerapan teknik dalam domain berbeda.

1.6 Sistematika Penulisan

Agar laporan penelitian lebih terstruktur dan mudah dipahami, peneliti Menyusun sistematika penulisan sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah yang mendasari penelitian, rumusan masalah yang hendak dijawab melalui penelitian, batasan masalah untuk memperjelas fokus penelitian, tujuan yang ingin dicapai, serta manfaat penelitian. Selain itu, bab ini juga mencakup sistematika penulisan untuk memberikan gambaran umum mengenai alur penelitian skripsi ini.

BAB II TINJAUAN PUSTAKA

Bab ini mencakup studi literatur yang memuat tinjauan atas penelitian sebelumnya terkait topik ini, serta teori-teori dasar yang digunakan sebagai landasan dalam penelitian. Teori yang dibahas mencakup konsep deteksi kecurangan, pengelolaan data tidak seimbang, serta algoritma machine learning seperti Random Forest, XGBoost, dan Support Vector Machine yang relevan dengan topik penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan objek penelitian, alur penelitian, serta alat dan bahan yang digunakan dalam penelitian. Proses seperti pengumpulan data, preprocessing, pemodelan menggunakan algoritma machine learning, serta evaluasi model juga dijelaskan secara rinci.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian, termasuk hasil deteksi kecurangan menggunakan algoritma yang telah diterapkan. Pembahasan dilakukan untuk menginterpretasikan hasil yang telah diperoleh dan menilai keefektifan model dalam mendeteksi kecurangan pada transaksi e-commerce. Analisis performa model juga disertakan untuk membandingkan metode yang digunakan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan, serta saran-saran yang dapat diberikan kepada pihak-pihak terkait. Saran ini mencakup pengembangan penelitian lebih lanjut maupun implementasi nyata di dunia e-commerce untuk meningkatkan keamanan transaksi dan mendeteksi potensi kecurangan secara efektif.

